

IN THE SPECIFICATION

Please delete the paragraph beginning at page 10, line 24 and ending at page 11, line 2 and substitute therefor the following new paragraph:

A1
An introduction to genetic algorithms can be found in David E. Goldberg (1989) Genetic Algorithms in Search, Optimization and Machine Learning Addison-Wesley Pub Co; ISBN: 0201157675 and in Timothy Masters (1993) Practical Neural Network Recipes in C++ (Book&Disk edition) Academic Pr; ISBN: 0124790402. A variety of more recent references discuss the use of genetic algorithms used to solve a variety of difficult problems. See, e.g., garage.cse.msu.edu/papers/papers-index.html (on the world wide web) and the references cited therein; gaslab.cs.unr.edu/ (on the world wide web) and the references cited therein; aic.nrl.navy.mil/ (on the world wide web) and the references cited therein; cs.gmu.edu/research/gag/ (on the world wide web) and the references cited therein and cs.gmu.edu/research/gag/pubs.html (on the world wide web) and the references cited therein.

Please delete the paragraph beginning at page 18, line 26 and ending at page 19, line 17 and substitute therefor the following new paragraph:

A2
One example algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (ncbi.nlm.nih.gov/ on the world wide web). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score

A²
goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

Please delete the paragraph beginning at page 22, line 13 and ending at page 22, line 26 and substitute therefor the following new paragraph:

A³
For example, oligonucleotides *e.g.*, for use in *in vitro* amplification/ gene reconstruction methods, for use as gene probes, or as shuffling targets (*e.g.*, synthetic genes or gene segments) are typically synthesized chemically according to the solid phase phosphoramidite triester method described by Beaucage and Caruthers (1981), *Tetrahedron Letts.*, 22(20):1859-1862, *e.g.*, using an automated synthesizer, as described in Needham-VanDevanter *et al.* (1984) *Nucleic Acids Res.*, 12:6159-6168. Oligonucleotides can also be custom made and ordered from a variety of commercial sources known to persons of skill. There are many commercial providers of oligo synthesis services, and thus this is a broadly accessible technology. Any nucleic acid can be custom ordered from any of a variety of commercial sources, such as The Midland Certified Reagent Company (mcr@oligos.com), The Great American Gene Company (genco.com, on the world wide web), ExpressGen Inc. (expressgen.com, on the world wide web), Operon Technologies Inc. (Alameda, CA) and many others. Similarly, peptides and antibodies can be custom ordered from any of a variety of sources, such as PeptidoGenic (pkim@ccnet.com), HTI Bio-products, inc. (htibio.com, on the world wide web), BMA Biomedicals Ltd (U.K.), Bio-Synthesis, Inc., and many others.

Please delete the paragraph beginning at page 44, line 28 and ending at page 45, line 8 and substitute therefor the following new paragraph:

A⁴
If the assay conditions are then altered in only one parameter, different individuals from the library will be identified as the best performers. Because the screening conditions are very similar, most amino acids are conserved between the two sets of best performers. Comparisons of the sequences (*e.g.*, *in silico*) of the best enzymes under the two different conditions identifies the sequence differences responsible for the differences in performance. Principal component analysis is a powerful tool to use for identifying sequences conferring a particular property. For example,

Partek Incorporated (St. Peters, Missouri; partek.com, on the world wide web) provides software for pattern recognition (e.g., which provide Partek Pro 2000 Pattern Recognition Software) which can be applied to genetic algorithms for multivariate data analysis, interactive visualization, variable selection, neural & statistical modeling. Relationships can be analyzed, e.g., by Principal Components Analysis (PCA) mapped scatterplots and biplots, Multi-Dimensional Scaling (MDS) mapped scatterplots, Star plots, etc.

Please delete the paragraph beginning at page 47, line 5 and ending at page 47, line 13 and substitute therefor the following new paragraph:

For example, neural net approaches can be coupled to genetic algorithm-type programming. for example, NNUGA (Neural Network Using Genetic Algorithms) is an available program (cs.bgu.ac.il/~omri/NNUGA/, on the world wide web) which couples neural networks and genetic algorithms. An introduction to neural networks can be found, e.g., in Kevin Gurney (1999) An Introduction to Neural Networks, UCL Press, 1 Gunpowder Square, London EC4A 3DE, UK. and at shef.ac.uk/psychology/gurney/notes/index.html (on the world wide web) additional useful neural network references include those noted above in regard to genetic algorithms and, e.g., Christopher M. Bishop (1995) Neural Networks for Pattern Recognition Oxford Univ Press; ISBN: 0198538642; Brian D. Ripley, N. L. Hjort (Contributor) (1995) Pattern Recognition and Neural Networks Cambridge Univ Pr (Short); ISBN: 0521460867.

Please delete the paragraph beginning at page 50, line 23 and ending at page 51, line 20 and substitute therefor the following new paragraph:

With respect to modeling and structural analysis, a reductionist approach, in which protein positions are classified by their local environments, has aided development of appropriate energy expressions (reviewed in Street and Mayo (1999) "Computational Protein Design" Structure with Folding and Design 7(5):R105-R109). For examples of cycles of modeling and structural analysis, See, e.g., mayo.caltech.edu/ (on the world wide web); Gordon and Mayo (1999) "Branch-and-Terminate: A Combinatorial Optimization Algorithm for Protein Design" Structure with Folding and Design 7(9):1089-1098; Street and Mayo (1999) "Intrinsic β -sheet Propensities Result from van der Waals Interactions Between Side Chains and the Local Backbone" Proc. Natl. Acad. Sci. USA, 96, 9074-9076; Gordon et al. (1999) "Energy Functions for Protein Design" Current Opinion in Structural Biology 9(4):509-513; Street and Mayo (1999) "Computational Protein Design" Structure with Folding and Design 7(5):R105-R109; Strop and Mayo (1999) "Rubredoxin

Variant Folds Without Iron" J. Am. Chem. Soc. 121(11):2341-2345; Gordon and Mayo (1998) "Radical Performance Enhancements for Combinatorial Optimization Algorithms based on the Dead-End Elimination Theorem" J. Comp. Chem 19:1505-1514; Malakauskas and Mayo (1998) "Design, Structure, and Stability of a Hyperthermophilic Protein Variant" Nature Struct. Biol. 5:470; Street and Mayo (1998) "Pairwise Calculation of Protein Solvent-Accessible Surface Areas" Folding & Design 3: 253-258. Dahiyat and Mayo (1997) "De Novo Protein Design: Fully Automated Sequence Selection" Science 278:82-87; Dahiyat and Mayo (1997) "Probing the Role of Packing Specificity in Protein Design" Proc. Natl. Acad. Sci. USA 94:10172-10177; Haney et al. (1997) "Structural basis for thermostability and identification of potential active site residues for adenylate kinases from the archaeal genus *Methanococcus*" Proteins 28(1):117-30; and, Dahiyat et al. (1997) "Automated Design of the Surface Positions of Protein Helices" Prot. Sci. 6:1333-1337; Dahiyat et al. (1997) "De Novo Protein Design: Towards Fully Automated Sequence Selection" J. Mol. Biol. 273:789-796. Protein design programs can be used to build or modify proteins with any selected set of design criteria and these design criteria can be used as filters for any GO as noted herein, and/or in recursive cycles of design (by modeling or structural analysis), in silico GO or GA application, and/or physical recombination of nucleic acids of interest.

Please delete the paragraph beginning at page 55, line 6 and ending at page 55, line 18 and substitute therefor the following new paragraph:

Similarly, molecular dynamic simulations such as those above and, e.g., Ornstein et al. (emsl.pnl.gov:2080/homes/tms/bms.html (on the world wide web); Curr Opin Struct Biol (1999) 9(4):509-13) provide for "rational" enzyme redesign by biomolecular modeling & simulation to find new enzymatic forms that would otherwise have a low probability of evolving biologically. For example, rational redesign of p450 cytochromes and alkane dehalogenase enzymes are a target of current rational design efforts. Any rationally designed protein (e.g., new p450 homologues or new alkaline dehydrogenase proteins) can be evolved by reverse translation and shuffling against either other designed proteins or against related natural homologous enzymes. Details on p450s can be found in Ortiz de Montellano (ed.) 1995, Cytochrome P450 Structure and Mechanism and Biochemistry, Second Edition Plenum Press (New York and London). Furthermore, the dynamic simulations can be used as predictors of residues of interest and coupling or linkages between residues of interest and cross-over selection (or other GOs) can be performed to specifically maintain or eliminate such linkages in silico.

Please delete the paragraph beginning at page 64, line 28 and ending at page 65, line 6 and substitute therefor the following new paragraph:

A⁸

HMM can be used in other ways as well. Instead of applying the generated profile to identify previously unidentified family members, the HMM profile can be used as a template to generate de novo family members (e.g., intermediate members of a cladistic tree of nucleic acids). For example, the program, HMMER is available (hmmer.wustl.edu/, on the world wide web). This program builds a HMM profile on a defined set of family members. A sub-program, HMMEMIT, reads the profile and constructs de novo sequences based on that. The original purpose of HMMEMIT is to generate positive controls for the search pattern, but the program can be adapted to the present invention by using the output as in silico generated progeny of a HMM profile defined shuffling. According to the present invention, oligonucleotides corresponding to these nucleic acids are generated for recombination, gene reconstruction and screening.

Please delete the paragraph beginning at page 73, line 3 and ending at page 73, line 23 and substitute therefor the following new paragraph:

A⁹

Typically, PDA starts with a protein backbone structure and designs the amino acid sequence to modify the protein's properties, while maintaining it's three dimensional folding properties. Large numbers of sequences can be manipulated using PDA, allowing for the design of protein structures (sequences, subsequences, etc.). PDA is described in a number of publications, including, e.g., Malakauskas and Mayo (1998) "Design, Structure and Stability of a Hyperthermophilic Protein Variant" Nature Struc. Biol. 5:470; Dahiyat and Mayo (1997) "De Novo Protein Design: Fully Automated Sequence Selection" Science, 278, 82-87. DeGrado, (1997) "Proteins from Scratch" Science, 278:80-81; Dahiyat, Sarisky and Mayo (1997) "De Novo Protein Design: Towards Fully Automated Sequence Selection" J. Mol. Biol. 273:789-796; Dahiyat and Mayo (1997) "Probing the Role of Packing Specificity in Protein Design" Proc. Natl. Acad. Sci. USA, 94:10172-10177; Hellinga (1997) "Rational Protein Design – Combining Theory and Experiment" Proc. Natl. Acad. Sci. USA, 94:10015-10017; Su and Mayo (1997) "Coupling Backbone Flexibility and Amino Acid Sequence Selection in Protein Design" Prot. Sci. 6:1701-1707; Dahiyat, Gordon and Mayo (1997) "Automated Design of the Surface Positions of Protein Helices" Prot. Sci., 6:1333-1337; Dahiyat and Mayo (1996) "Protein Design Automation" Prot. Sci., 5:895-903. Additional details regarding PDA are available, e.g., at xencor.com (on the world wide web).

Please delete the paragraph beginning at page 80, line 27 and ending at page 81, line 2 and substitute therefor the following new paragraph:

A10
3. Potential sequences for recombination can also be identified by computational methods other than by direct homology or structural information, e.g. by the PRINTS system (Attwood TK, Beck ME, Bleasby AJ, Degtyarenko K, Parry Smith DJ *Nucleic Acids Res* 1996 Jan 1;24(1):182-8), Shotgun (Pegg and Babbitt (1999) *Bioinformatics* 15(9):729-40) independently evolving sequence modules (may correspond to folding units as in MasterCatalog (eragen.com, on the world wide web) and other sequence comparison methods.

Please delete the paragraph beginning at page 81, line 9 and ending at page 81, line 17 and substitute therefor the following new paragraph:

A11
Similarly, PRINTS (e.g., Atwood et al., *above*) is a compendium of protein motif fingerprints derived from the OWL composite sequence database. Fingerprints are groups of motifs within sequence alignments whose conserved nature allows them to be used as signatures of family membership. Fingerprints can provide improved diagnostic reliability over single motif methods by virtue of the mutual context provided by motif neighbors. The database is now accessible via the UCL Bioinformatics Server on biochem.ucl.ac.uk/bsm/dbbrowser/ (on the world wide web). Atwood et al. describe the database, its compilation and interrogation software, and its Web interface. *See also*, Attwood et al. (1997) "Novel developments with the PRINTS protein fingerprint database" *Nucleic Acids Res* 25(1):212-7.

Please delete the paragraph beginning at page 88, line 22 and ending at page 89, line 3 and substitute therefor the following new paragraph:

A12
One approach to screening diverse libraries is to use a massively parallel solid-phase procedure to screen cells expressing shuffled nucleic acids, e.g., which encode enzymes for enhanced activity. Massively parallel solid-phase screening apparatus using absorption, fluorescence, or FRET are available. *See*, e.g., United States Patent 5,914,245 to Bylina, et al. (1999); *see also*, kairos-scientific.com/ (on the world wide web); Youvan et al. (1999) "Fluorescence Imaging Micro-Spectrophotometer (FIMS)" *Biotechnology et alia* <et-al.com (on the world wide web)> 1:1-16; Yang et al. (1998) "High Resolution Imaging Microscope (HIRIM)" *Biotechnology et alia*, <et-al.com (on the world wide web)> 4:1-20; and Youvan et al. (1999) "Calibration of Fluorescence Resonance Energy Transfer in Microscopy Using Genetically Engineered GFP Derivatives on Nickel Chelating Beads" posted at kairos-scientific.com (on the world wide web).